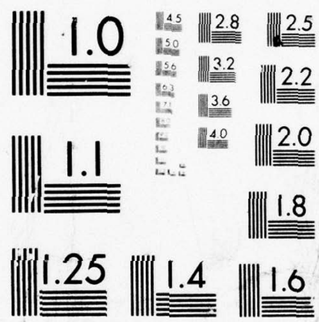
AD-A052 543 NATIONAL RANGE OPERATIONS DIRECTORATE WHITE SANDS MIS--ETC F/G 12/1
APPLICATION OF ROBUST STATISTICAL METHODS TO DATA REDUCTION.(U)
MAR 78 W S AGEE, R H TURNER

UNCLASSIFIED ACD-TR-65 NL

| OF |

AD
A052543

END
DATE
FILMED
5-78
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

TECHNICAL REPORT

NO. 65

MARCH 1978

APPLICATION OF ROBUST STATISTICAL METHODS TO DATA REDUCTION

MATHEMATICAL SERVICES BRANCH
ANALYSIS & COMPUTATION DIVISION
US ARMY WHITE SANDS MISSILE RANGE
WHITE SANDS MISSILE RANGE, NEW MEXICO

TECHNICAL REPORT

NO. 65

Prepared by William S Agee
WILLIAM S. AGEE
Mathematician, Math Svcs Br

Robert H. Turner
ROBERT H. TURNER
Mathematician, Math Svcs Br

Reviewed by Jon E Gibson
JON E. GIBSON
Chief, Math Svcs Br

Approved by Patrick J. Higgins
PATRICK J. HIGGINS
Chief, Anal & Cmpt Div

ABSTRACT

Robust Statistics provides a fresh approach to the difficult problem of editing in data reduction. Of prime concern are grossly erroneous measurements which, when undetected, completely destroy automated data reduction procedures causing costly reruns and time delays with human detection of the erroneous measurements. The application of robust statistical methods has been highly successful in dealing with this problem. An introduction to the robust M-estimates and their numerical computation is given. The application of M-estimates to data preprocessing, instrument calibration, N-station cinetheodolites, N-station radar solution, and filtering are described in detail. Numerical examples of these applications using real measurements are given.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Analysis & Computation Division
National Range Operations Directorate
White Sands Missile Range, New Mexico 88002

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP
NA

3. REPORT TITLE
6
APPLICATION OF ROBUST STATISTICAL METHODS TO DATA REDUCTION.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
9 Technical rept.,

5. AUTHOR(S) (First name, middle initial, last name)
10
William S. Agee [redacted] Robert H. Turner

6. REPORT DATE
11
MAR[redacted]78

7a. TOTAL NO. OF PAGES
23

7b. NO. OF REFS
12

8a. CONTRACT OR GRANT NO.

b. PROJECT NO.

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
Technical Report No. 65
14 ACD-TR-65

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
12 26p.

10. DISTRIBUTION STATEMENT
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

Robust Statistics provides a fresh approach to the difficult problem of editing in data reduction. Of prime concern are grossly erroneous measurements which, when undetected, completely destroy automated data reduction procedures causing costly reruns and time delays with human detection of the erroneous measurements. The application of robust statistical methods has been highly successful in dealing with this problem. An introduction to the robust M-estimates and their numerical computation is given. The application of M-estimates to data preprocessing, instrument calibration, N-station cinetheodolites, N-station radar solution, and filtering are described in detail. Numerical examples of these applications using real measurements are given.

407 706

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

UNCLASSIFIED
Security Classification

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Robust						
Regression						
M-estimates						
Outliers						
Data Reduction						

UNCLASSIFIED
Security Classification

INTRODUCTION

Robust statistics provides a new approach to data editing in trajectory data reduction. Data editing, whose function is to deal with wild observations, has been a most frustrating problem for the data analyst. The use of robust statistics has been highly successful, much more so than previous methods, in dealing with this problem. There are several applications of robust statistics to data editing in trajectory data reduction. The applications considered here are:

- Data Preprocessing
- Instrument calibration
- N-station Cine solution
- N-station radar solution
- Filtering

Before describing these applications we need to answer: What are robust statistics and how do robust statistics apply to data editing? In answer to the first part of the question robust statistical methods are those which tend to retain their desirable properties under at least mild violations of the assumptions under which they were derived. Possibly a more useful description of a robust statistical procedure is one which will perform well under a variety of underlying distribution functions or in the presence of observations from contaminating distributions. Thus, the sample median is a more robust procedure than the sample mean for estimating the mean of a symmetric parent distribution, if even a moderate amount of contamination from a long tailed distribution is present. In answer to the second part of the question we are probably not very concerned about the performance of data reduction procedures under a wide variety of underlying distribution functions of the observations but are mainly concerned about the performance of our methods in the presence of observations from contaminating distributions, i.e., outliers. Thus, in data reduction we are interested in the development of robust statistical methods which are highly outlier resistant. In data reduction we are usually interested in estimating the parameters in some postulated linear or nonlinear model of a process. Thus, in data reduction we are specifically interested in developing methods for linear and nonlinear regression which are insensitive to a large percentage of outlying observations. Many sources of outliers are present in trajectory measuring systems. Without going into any detail, these sources may be broadly grouped into the categories of equipment malfunction, outside interference, and human error.

The usual methods of least squares, optimally weighted least squares, maximum likelihood, etc., used in data reduction for estimating parameters in a regression model are rendered useless by the presence of outliers. To quote Huber [1], "even a single grossly outlying observation may spoil the least squares estimate and moreover outliers are much harder to spot in the regression case than in the simple location case."

Although the history of robust estimation goes back to the 19th century, the development of robust regression methods is just currently becoming a popular topic for statistical research. Some of the earliest methods for robust regression were developed in the 1950's, notably the methods reported by Brown and Mood [9] and by Theil [5]. Robust estimation methods have been classified by Huber [1] and [2]. Huber's classifications are termed L-estimates, M-estimates, and R-estimates. The L-estimates are estimates which are linear combinations of the order statistics. The α-trimmed mean is an example of an L-estimate for a simple location parameter. The R-estimates are estimates derived on the basis or rank tests. The estimate of location obtained by taking the median of all pairwise averages of the observations is an R-estimate. Of the three classifications for robust estimates given by Huber we shall only be concerned with M-estimates in this report. The reason for this is not that M-estimates are superior but because we are only interested in describing the applications of robust regression to data reduction and this seemed easiest to do with the M-estimates.

Given the linear model

$$y_i = \sum_{j=1}^{P} x_{ij}\Theta_j + e_i \qquad i=1, N \tag{1}$$

the regression parameters Θ_i are to be estimated. The M-estimates of Θ_i minimize

$$\sum_{i=1}^{N} \rho(y_i - \sum_j x_{ij}\Theta_j) \tag{2}$$

where ρ $(\cdot)$ is some function which is often convex. Differentiating with respect to Θ leads to

$$\sum_{i=1}^{N} X_i^T \psi(y_i - \sum_j x_{ij}\Theta_j) = 0 \tag{3}$$

where

$$X_i^T = \mathrm{col}(x_{i1}, x_{i2}, - - - x_{ip})$$

and $\psi(\cdot)$ is the derivative of $\rho(\cdot)$. (3) is the analog of the normal equations in least squares regression. The estimate which results from solving (3) is called an M-estimate. Rather than specifying the function ρ, M-estimates are usually described by specifying the function ψ. If $f(y;\Theta)$ is the probability density function underlying the observations, and if

$$\psi = \frac{\partial f(y;\Theta)}{\partial \Theta} \Big/ f(y;\Theta)$$

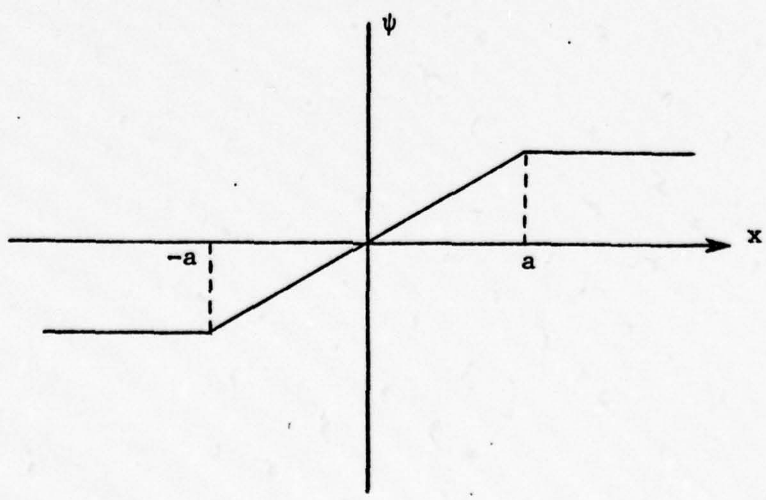
then the M-estimate obtained is the maximum likelihood estimate. Since the function ρ is usually not homogeneous, as it would be in least squares, the M-estimates obtained would usually not be scale invariant. Hence, to force scale invariance we minimize

$$\sum_{i=1}^{N} \rho\left(\frac{y_i - \sum_j x_{ij}\Theta_j}{s}\right) \tag{4}$$

where s is some measure of dispersion of the residuals, $y_i - \sum_j x_{ij}\Theta_j$. The measure s also needs to be a robust measure of dispersion.

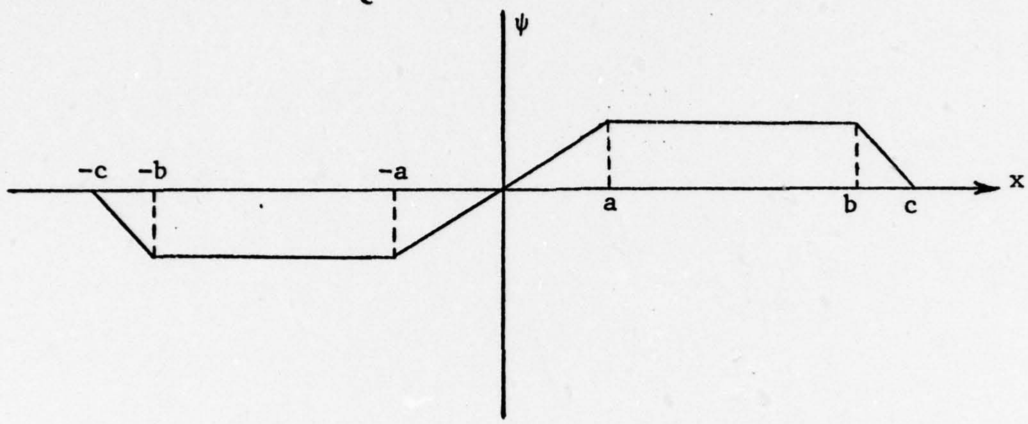
Several ψ functions have been proposed in the literature. Basically those ψ functions fall into two classes, the redescending type and non-redescending type. We will only consider one member of each of these classes in this report. The original ψ function proposed by Huber is of the non-redescending type. This function is

$$\psi(x) = \begin{cases} x & |x| \leq a \\ a\ \mathrm{sgn}(x) & |x| \geq a \end{cases} \tag{5}$$



An example of a ψ function of the redescending type is the function proposed by Hampel [6].

$$\psi(x) = \begin{cases} x & |x| \leq a \\ a\,\mathrm{sgn}(x) & a \leq |x| \leq b \\ a\left(\dfrac{x - c\,\mathrm{sgn}(x)}{b - c}\right) & b \leq |x| \leq c \\ 0 & |x| \geq c \end{cases} \tag{6}$$



Other ψ functions have been proposed by Andrews [8], Tukey [3], and Ramsay [4]. There are also a number of other methods for robust regression, see [11].

An attractive feature of least squares regression is the ease of numerical solution. One might be inclined to think that the numerical solution for an M-estimate would in many cases be prohibitive. At worst (4) can be minimized by one of the many algorithms for minimization, e.g., the Fletcher-Powell method [10].

An iterative solution such as a Gauss-Newton method can easily be applied to minimize (4). Setting the derivative of (4) to zero and linearizing about an arbitrary point $\hat{\Theta}^{\{\kappa\}}$ in the iteration sequence gives

$$\sum_{i=1}^{N} X_i^T \left[\psi\left(\frac{y_i - X_i\hat{\Theta}^{\{\kappa\}}}{s}\right) - \psi'\left(\frac{y_i - X_i\hat{\Theta}^{\{\kappa\}}}{s}\right)\frac{X_i}{s}\left(\hat{\Theta}^{\{\kappa+1\}} - \hat{\Theta}^{\{\kappa\}}\right)\right] = 0 \tag{7}$$

Solving (7) for $\hat{\Theta}^{\{\kappa+1\}} - \hat{\Theta}^{\{\kappa\}}$ yields

$$\hat{\Theta}^{\{\kappa+1\}} - \hat{\Theta}^{\{\kappa\}} = M^{-1} \sum_{i=1}^{N} \psi\left(\frac{y_i - X_i\hat{\Theta}^{\{\kappa\}}}{s}\right) X_i^T \tag{8}$$

where

$$M = \sum_{i=1} \psi'\left(\frac{y_i - X_i\hat{\Theta}^{\{\kappa\}}}{s}\right)\frac{X_i^T X_i}{s} \tag{9}$$

The iterative application of (8) and (9) results in a fairly simple method for obtaining an M-estimate. An approximate sample covariance for this estimate is given by

$$\mathrm{cov}(\hat{\Theta}) = V = \frac{1}{n-p}\sum_{i=1}^{N} \psi^2\left(\frac{y_i - X_i\hat{\Theta}}{s}\right) M^{-1}\left(\sum_{j=1}^{N} X_j^T X_j\right) M^{-1} \tag{10}$$

An even simplier numerical method and one which has achieved considerable popularity for obtaining M-estimates is the iterative application of weighted least squares. Setting the derivative of (4) with respect to Θ equal to zero gives

$$\sum_{i=1}^{N} X_i^T \psi\left(\frac{y_i - X_i\hat{\Theta}}{s}\right) = 0 \tag{11}$$

Now rewrite (11) as

$$\sum_{i=1}^{N} \frac{\psi\left(\frac{y_i - X_i\hat{\Theta}}{s}\right)}{\left(\frac{y_i - X_i\hat{\Theta}}{s}\right)} X_i^T(y_i - X_i\hat{\Theta}) = 0 \tag{12}$$

Let

$$W_i(\hat{\Theta}) = \frac{\psi\left(\frac{y_i - X_i\hat{\Theta}}{s}\right)}{\left(\frac{y_i - X_i\hat{\Theta}}{s}\right)} \tag{13}$$

Then (12) is

$$\sum_{i=1}^{N} W_i(\hat{\Theta})\, X_i^T(y_i - X_i\hat{\Theta}) = 0 \tag{14}$$

(14) can be solved iteratively as follows. Starting at an arbitrary point $\hat{\Theta}^{\{\kappa\}}$ in the sequence of iterations, we replace (14) by

$$\sum_{i=1}^{N} W_i(\hat{\Theta}^{\{\kappa\}})X_i^T(y_i - X_i\hat{\Theta}^{\{\kappa + 1\}}) = 0 \tag{15}$$

Solving (15) for $\hat{\Theta}^{\{\kappa + 1\}}$

$$\hat{\Theta}^{\{\kappa + 1\}} = \left(\sum_{j=1}^{N} W_j(\hat{\Theta}^{\{\kappa\}})X_j^T X_j\right)^{-1} \sum_{i=1}^{N} W_i(\hat{\Theta}^{\{\kappa\}})X_i^T y_i \tag{16}$$

Thus, we can use an ordinary weighted least squares algorithm iteratively to obtain the M-estimate.

Throughout the discussion of M-estimates we have used the dispersion measure s of the residuals $y_i - X_i\Theta$ without consideration for its computation. Robust dispersion measures are often taken to be a multiple of the inter-quartile range or of some other range statistic of a set of residuals. A dispersion measure which seems to be most popular with those using M-estimates is the median deviation or the MAD (Median of the Absolute Deviations) estimate as it is sometimes called. The MAD estimate used above is defined by

$$s = \operatorname*{med}_{i}|r_i| \Big/ .6745 \tag{17}$$

where $r_i = y_i - X_i\Theta$. Hampel [6] has shown that the MAD estimate is the most robust estimate of dispersion.

Both the Gauss-Newton method and the weighted least squares method for obtaining M-estimates are iterative and therefore require a starting solution. The required closeness of the starting solution to the final solution is dependent on the application and the type of ψ function used. Quite often an ordinary unweighted least squares solution is a sufficiently good starting solution. In some cases it is necessary to use a starting solution which is more robust, see [12].

APPLICATION TO DATA PREPROCESSING

It is this application which provided our original motivation for the development and application of robust statistical methods in data reduction. There are several possible functions of data preprocessing. Ambiguities in phase measurements might be resolved by preprocessing. It might be used merely to detect outliers in the measurement data because their detection in the main processor might be considerably more difficult. Also, the main processor often requires the use of weights for each of the measurements or the main processor might require that a set of measurements be synchronized before processing. These requirements can be fulfilled by data preprocessing.

Given the time history of a particular measurement function for its entire span of observation on a trajectory, the preprocessing function divides the interval of observation into equal segments of T seconds except for a final segment either shorter or longer that T. Over each of these segments a polynomial, usually a quadratic, is fit to the measurements. Alternatively, a cubic spline might be fit to the entire span of measurement data using the end points of the T second segments as knot times. Thus, in measurement preprocessing we might model the αth measurement over an arbitrary interval of the trajectory as

$$y_\alpha(t_i) = \Theta_0 + \Theta_1(t_i - \bar{t}) + \Theta_2(t_i - \bar{t})^2 \qquad i=1, N \tag{18}$$

where

$$\bar{t} = \frac{1}{N}\sum_{i=1}^{N} t_i$$

Using some robust M-estimate of the parameter vector $\Theta = [\Theta_0\ \Theta_1\ \Theta_2]$ we would minimize

$$Q = \sum_{i=1}^{N} \rho\left(\frac{y_\alpha(t_i) - \Theta_0 - \Theta_1(t_i - \bar{t}) - \Theta_2(t_i - \bar{t})^2}{s}\right) \tag{19}$$

which upon differentiating gives the analog of the normal equations

$$\sum_{i=1}^{N} T_i^T \psi\left(\frac{y_\alpha(t_i) - \hat{\Theta}_0 - \hat{\Theta}_1(t_i - \bar{t}) - \hat{\Theta}_2(t_i - \bar{t})^2}{s}\right) \tag{20}$$

where

$$T_i = \left[1\ (t_i - \bar{t})\ (t_i - \bar{t})^2\right]$$

We solve (20) iteratively to obtain robust estimates $\hat{\Theta}_0$, $\hat{\Theta}_1$, $\hat{\Theta}_2$.

In the iterative solution of (20) s is taken to be the median of absolute residuals

$$s = \operatorname*{med}_i |r_\alpha(t_i)| \Big/ .6745$$

where

$$r_\alpha(t_i) = y_\alpha(t_i) - \hat{\Theta}_0 - \hat{\Theta}_1(t_i - \bar{t}) - \hat{\Theta}_2(t_i - \bar{t})^2$$

The following data set is from a real data reduction situation. The measurements are a sequence of azimuth measurements from a cine.

OBSERVATIONS	RESIDUALS FROM ROBUST FIT	RESIDUALS FROM LEAST SQUARES FIT
-1.70987	.000012	-.157774
-1.70942	-.000004	-.000204
-1.70893	.000003	.105480
-1.70845	-.000015	.159227
-1.70793	-.000010	.161087
-1.70741	-.000021	.111021
-1.70682	.000022	.009099
-1.70626	.000019	-.144780
-1.70571	-.000010	-.350595
-1.70510	.000005	-.608277
-1.70449	.000004	-.917885
1.43777	3.141637	1.862231
1.44602	3.149243	1.456410
-1.70257	-.000007	-2.158177
1.44667	3.146558	.473139

There are three obvious outliers in the data. The residuals from an ordinary least square fit, which are given in the last column yield no information about outliers in the data. The residuals from the robust fit which were obtained using a Hampel ψ function (breakpoints 2.5, 5.0, 7.5) show exactly which observations were outliers. Outliers can be detected as those residuals r_i for which $r_i \geq ks$. The dispersion s may be saved for use in making weights for the observations in the main processing. Another example of data preprocessing is provided by the 40 point data sequence below.

	LEAST SQUARES RESIDUALS	ROBUST RESIDUALS	OBSERVATION	NORMALIZED ROBUST RESIDUALS
1	-.011022	-.000278	.20642275	1.005559
2	-.009071	-.000006	.20973521	.020803
3	-.007471	-.000033	.21296912	.120171
4	-.005711	.000151	.21663652	.546808
5	-.004461	-.000123	.22006619	.445501
6	-.002730	.000136	.22425138	.492246
7	-.001590	-.000144	.22811853	.519552
8	-.000213	-.000135	.23249603	.487267
9	.001201	-.000038	.23718297	.136926
10	.002489	-.000014	.24201791	.051970
11	.003798	.000082	.24714760	.297949
12	.005624	.000748	.25306741	2.703007
13	.005421	-.000564	.25723122	2.037977
14	.008660	.001617	.26510980	5.845255
15	.006010	-.002037	.26737381	7.361710
16	.009663	.000662	.27621340	2.394020
17	.011016	.001113	.28302583	4.023731
18	.010359	-.000392	.28810282	1.418292
19	.011568	.000019	.29531815	.067036
20	.012005	-.000291	.30203451	1.051051
21	.012861	-.000129	.30944403	.464413
22	.013557	-.000075	.31696650	.269818
23	.014001	-.000222	.32450959	.800901
24	.014501	-.000260	.33238295	.938668
25	.015039	-.000209	.34056693	.754131
26	.015433	-.000249	.34888132	.898506
27	.015913	-.000152	.35755414	.547835
28	.016283	-.000113	.36639033	.406971
29	.016494	-.000181	.37534057	.654202
30	-.058265	-.075167	.30959446	271.639732
31	-.172487	-.189565	.20465789	685.052254
32	.018472	.001270	.40517605	4.589248
33	-.064416	-.081690	.33212063	295.211357
34	.089274	.071980	.49591643	260.122231
35	-.251831	-.269092	.16519139	972.446930
36	.007852	-.009326	.43552655	33.701152
37	.159606	.142564	.59820610	515.197899
38	.059168	.042313	.50896735	152.912771
39	.016704	.000088	.47797510	.318960
40	.016296	-.000028	.48931307	.101770

The solution for the M-estimate used a least square starting solution and a Hampel ψ function with breakpoints at 2.5, 5, and 7.5. In the list of least square residuals given above some of the outliers are obvious while others are not. The column of normalized residuals is merely the robust residual divided by the robust dispersion measure s. If we declare that residuals greater than 2.5s are outliers then we would flag observations 12, 14, 15, 17, 30, 31, 32, 33, 34, 35, 36, 37, and 38 as outliers. Some of these outliers are much more gross than others. The M-estimate of the parameter vector is $\hat{\Theta}_0 = .20388$, $\hat{\Theta}_1 = .05419$, $\hat{\Theta}_2 = .04427$. This example is simulated data so that the true parameter vector is known to be $\Theta_0 = .20397$, $\Theta_1 = .0537$, $\Theta_2 = .0445$. The least squares starting solution was $\Theta^{\{o\}}_0 = .21636$, $\Theta^{\{o\}}_1 = .01901$, $\Theta^{\{o\}}_2 = .05466$.

INSTRUMENT CALIBRATION

Surveyed targets are used for calibrating, i.e., estimating the coefficients in an error model, for radars, cinetheodolites or laser trackers. Suppose for example we have M surveyed targets for a laser tracker. Let R_{sj}, A_{sj}, E_{sj} be the surveyed range, azimuth, and elevation for the jth target. Suppose that multiple observations of the targets are available so that we have N_j observations for the jth target. Denote these range, azimuth and elevation observations by R_{ij}, A_{ij} and E_{ij}, $i = 1, N_j$, $j = 1, M$. Let

$$\Delta R_{ij} = R_{ij} - R_{sj} = r_j^T\Theta + \text{(random error)}$$

$$\Delta A_{ij} = A_{ij} - A_{sj} = a_j^T\Theta + \text{(random error)}$$

$$\Delta E_{ij} = E_{ij} - E_{sj} = e_j^T\Theta + \text{(random error)}$$

where Θ is an unknown parameter vector and r_j, a_j, and e_j are known vectors. A common model for r_j, a_j and e_j is given by

$$r_j^T\Theta = \Theta_1 + \Theta_2 R_{sj} \tag{21}$$

$$a_j^T\Theta = \Theta_3 - \Theta_4 \tan E_{sj} \cos A_{sj} - \Theta_5 \tan E_{sj} \sin A_{sj} - \Theta_6/\cos E_{sj} \tag{22}$$

$$e_j^T\Theta = \Theta_7 + \Theta_4 \sin A_{sj} - \Theta_5 \cos A_{sj} \tag{23}$$

The usual least squares estimate of the parameter vector Θ would minimize

$$\sum_{j=1}^{M} \sum_{i=1}^{N_j} \left[(\Delta R_{ij} - r_j^T\Theta)^2 + (\Delta A_{ij} - a_j^T\Theta)^2 + (\Delta E_{ij} - e_j^T\Theta)^2 \right] \tag{24}$$

An M-estimate alternative to least squares would minimize

$$\sum_{j=1}^{M} \sum_{i=1}^{N_j} \left[\rho\left(\frac{\Delta R_{ij} - r_j^T\Theta}{s_r}\right) + \rho\left(\frac{\Delta A_{ij} - a_j^T\Theta}{s_a}\right) + \rho\left(\frac{\Delta E_{ij} - e_j^T\Theta}{s_e}\right) \right] \tag{25}$$

Differentiating (25) gives the analog to the normal equations

$$\sum_{j=1}^{M} \sum_{i=1}^{N_j} \left[\psi\left(\frac{\Delta R_{ij} - r_j^T\hat{\Theta}}{s_r}\right) \frac{r_j}{s_r} + \psi\left(\frac{\Delta A_{ij} - a_j^T\hat{\Theta}}{s_a}\right) \frac{a_j}{s_a} + \psi\left(\frac{\Delta E_{ij} - e_j^T\hat{\Theta}}{s_e}\right) \frac{e_j}{s_e} \right] = 0 \tag{26}$$

An iterative solution of (26) with

$$s_r = \operatorname*{med}_{i,j} |d_r(i,j)| \Big/ .6745$$

$$s_a = \operatorname*{med}_{i,j} |d_a(i,j)| \Big/ .6745$$

$$s_e = \operatorname*{med}_{i,j} |d_e(i,j)| \Big/ .6745$$

Where

$$d_r(i,j) = \Delta R_{ij} - r_j^T\hat{\Theta} \tag{27}$$

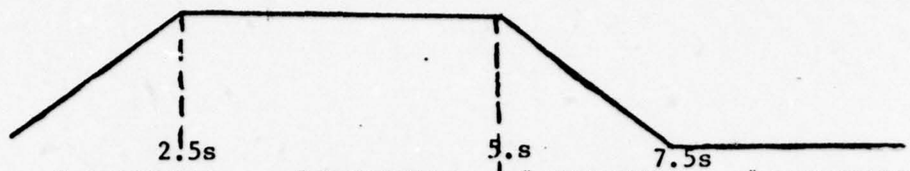
$$d_a(i,j) = \Delta A_{ij} - a_j^T\hat{\Theta} \tag{28}$$

$$d_e(i,j) = \Delta E_{ij} - e_j^T\hat{\Theta} \tag{29}$$

gives a robust estimate $\hat{\Theta}$. Since the elements of the parameter vector are usually small, the elements of the starting solution $\Theta^{\{o\}}$ may be set to zero except for $\Theta_1^{\{o\}}$, $\Theta_3^{\{o\}}$, and $\Theta_7^{\{o\}}$ which can be set to the medians of R_{ij}, A_{ij}, and E_{ij}, respectively.

The following example illustrates the application of M-estimates to the calibration of a laser tracker using real field data. The laser tracker is calibrated by using azimuth and elevation observations from eight reflective targets arranged in a circular pattern around the tracker at a range of about 2500 feet. We use the error model given in (21) - (23). Since the elevations of the eight calibration targets are approximately equal, it is obviously impossible to estimate Θ_6 in (22) without additional observations. In order to provide these additional observations we observe the same calibration targets but with the tracker "dumped", i.e., with an azimuth of approximately $A_{si} + 180^o$ and an elevation of approximately $E_{si} - 180^o$. These additional observations are called dumped readings and are treated as additional calibration targets. Also, we can see from (21) that we will be unable to estimate Θ_2 using the eight calibration targets since the ranges to all targets are approximately equal. In order to estimate Θ_2 we observe four additional calibration targets with ranges varying from 20000 feet to 60000 feet. In this example dumped readings were not available so that Θ_6 could not be estimated. Also, data from two of the close targets are missing. Approximately 250 observations are available for each of the remaining target boards.

A Hampel ψ function which was defined in (6) was used for this example. The parameters or break points of the Hampel ψ in this example are a = 2.5, b = 5.0, c = 7.5. The results of this robust calibration are summarized in the figure below by tabulating the number of residuals for each target lying in each region of the Hampel ψ. We show only the positive side of the ψ function with the number of residuals in each region being the sum of the numbers of residuals in the positive and corresponding negative side of the ψ function.



TARGET #	# RESIDUALS	# RESIDUALS	# RESIDUALS	# RESIDUALS
1	250	0	0	0
2	250	0	0	0
3	246	0	0	0
4	250	0	0	0
5	250	0	0	0
6	250	0	0	0
9	245	4	0	1
10	83	141	26	0
11	215	32	2	1
12	0	0	0	250

DISTRIBUTION OF RANGE RESIDUALS

TARGET #	# RESIDUALS	# RESIDUALS	# RESIDUALS	# RESIDUALS
1	250	0	0	0
2	250	0	0	0
3	246	0	0	0
4	250	0	0	0
5	250	0	0	0
6	250	0	0	0
9	245	5	0	0
10	238	12	0	0
11	247	3	0	0
12	248	2	0	0

DISTRIBUTION OF AZIMUTH RESIDUALS

TARGET #	# RESIDUALS	# RESIDUALS	# RESIDUALS	# RESIDUALS
1	250	0	0	0
2	250	0	0	0
3	246	0	0	0
4	250	0	0	0
5	250	0	0	0
6	250	0	0	0
9	245	5	0	0
10	238	12	0	0
11	247	3	0	0
12	248	2	0	0

DISTRIBUTION OF ELEVATION RESIDUALS

Thus, we can see that a significant percentage of the range observations from the long range targets are bad including all range observations from target number 12. The parameter estimates for this example are Θ_1 = 1.69 feet, $\Theta_2 = .508145 \times 10^{-4}$, Θ_3 = .11405 mr Θ_4 = .511565 mr, Θ_5 = −.105925 mr, Θ_7 = −.04014 mr. The least squares calibration for this example gives the erroneous values for the range calibration Θ_1 = −395.27 feet and $\Theta_2 = -.96948 \times 10^{-2}$.

N-STATION CINETHEODOLITE SOLUTION

The N-station Cine solution is a standard problem in data reduction. In this situation we are given azimuth observations $a_\alpha(t_i)$ and elevation observations $e_\alpha(t_i)$, $\alpha = 1, N_i$, from N_i cines at each time point t_i along a trajectory. From these N_i cines we must estimate the cartesian positions $x(t_i)$, $y(t_i)$, $z(t_i)$, at each time point. The observations are $a_\alpha(t_i) = A_\alpha(\bar{x}_i)$ + error and $e_\alpha(t_i) = E_\alpha(\bar{x}_i)$ + error. The measurement functions $A_\alpha(\bar{x}_i)$ and $E_\alpha(\bar{x}_i)$ are functions of the position vector $\bar{x}_i = [x(t_i)\ y(t_i)\ z(t_i)]$. These measurement functions are given by

$$A_\alpha(\bar{x}_i) = \tan^{-1} \frac{x(t_i) - x_\alpha}{y(t_i) - y_\alpha} \tag{30}$$

$$E_\alpha(x) = \tan^{-1} \frac{z(t_i) - z_\alpha}{\left[(x(t_i) - x_\alpha)^2 + (y(t_i) - y_\alpha)^2\right]^{1/2}} \tag{31}$$

where $(x_\alpha, y_\alpha, z_\alpha)$ is the cartesian position of the α^{th} cine. The usual least square problem to estimate the position $x(t_i)$, $y(t_i)$, $z(t_i)$ is nonlinear. Thus, the robust estimation of these quantities will be nonlinear both because the objective function for the robust estimation problem is non-quadratic and because the measurement model is a nonlinear function of the parameters to be estimated. The usual least squares solution would minimize

$$\sum_{\alpha=1}^{N_i} \left[\left(a_\alpha(t_i) - A_\alpha(\bar{x}_i)\right)^2 \cos^2 e_\alpha(t_i) + \left(e_\alpha(t_i) - E_\alpha(\bar{x}_i)\right)^2 \right] \tag{32}$$

An M-estimate of the position vector $\bar{x}_i$ would minimize

$$\sum_{\alpha=1}^{N_i} \left[\rho\left(\frac{a_\alpha(t_i) - A_\alpha(\bar{x}_i)}{s_a}\right) \cos^2 e_\alpha(t_i) + \rho\left(\frac{e_\alpha(t_i) - E_\alpha(\bar{x}_i)}{s_e}\right)\right] \tag{33}$$

Differentiating (33) gives

$$\sum_{\alpha=1}^{N_i} \left[\frac{1}{s_a} \psi\left(\frac{r_a(\alpha)}{s_a}\right) \frac{\partial A_\alpha(\bar{x}_i)}{\partial \bar{x}_i} \cos^2 e_\alpha(t_i) + \frac{1}{s_e} \psi\left(\frac{r_e(\alpha)}{s_e}\right) \frac{\partial E_\alpha(\bar{x}_i)}{\partial \bar{x}_i}\right] = 0 \tag{34}$$

where

$$r_a(\alpha) = a_\alpha(t_i) - A_\alpha(\bar{x}_i)$$

$$r_e(\alpha) = e_\alpha(t_i) - E_\alpha(\bar{x}_i)$$

(34) can be rewritten as

$$\sum_{\alpha=1}^{N_i} C_\alpha^T(\bar{x}_i)\, \psi\left(\frac{r_a(\alpha)}{s_a}, \frac{r_e(\alpha)}{s_e}\right) = 0 \tag{35}$$

where

$$C_\alpha^T(\bar{x}_i) = \left[\frac{1}{s_a} \frac{\partial A_\alpha(\bar{x}_i)}{\partial \bar{x}_i} \;\; \frac{1}{s_e} \frac{\partial E_\alpha(\bar{x}_i)}{\partial \bar{x}_i} \right]$$

a 3 x 2 matrix and

$$\psi\left(\frac{r_a(\alpha)}{s_a}, \frac{r_e(\alpha)}{s_e}\right) = \begin{bmatrix} \psi\left(\frac{r_a(\alpha)}{s_a}\right) \cos^2 e_\alpha(t_i) \\ \psi\left(\frac{r_e(\alpha)}{s_e}\right) \end{bmatrix} \tag{36}$$

An iterative solution of (35) with $s_a = \underset{\alpha}{\text{med}}|r_a(\alpha)|/.6745$, $s_e = \underset{\alpha}{\text{med}}|r_e(\alpha)|/.6745$ gives a robust estimate of the parameter vector x_1.

As an example or robust estimation applied to a cinetheodolite solution consider the following situation which is rather extreme but sometimes occurs. A missile is fired at a drone and cinetheodolites are observing both the missile and drone. It is required to provide a cine derived trajectory on both the missile and the drone. Due to an inadvertent clerical error one of the cines which was actually observing the missile was erroneously listed as observing the drone. Obviously, when doing a least squares solution to obtain the drone trajectory, the azimuth and elevations from one cine will be gross outliers and may destroy the least squares solution for the drone position coordinates. A single point example of this situation is furnished by the actual cine data given below

Cine	Obs. Azimuth	Obs. Elevation
1	.568106	.338886
2	-.626010	.122620
3	-2.665036	.359168
4	1.926249	.327177

Cine 2 is the one which is actually tracking the missile rather than the drone. Obviously, as in most situations which are the nonlinear, there is no way of distinguishing the outliers by inspecting the observations. As always in robust estimation a preliminary solution is required to start the iteration. Let $(x_\alpha, y_\alpha, z_\alpha)$ be a position solution obtained from the α^{th} pair of cines. In this example we have six possible pairs of cines so that $\alpha = 1, 6$. We then start the iteration with (x^o, y^o, z^o) where $x^o = \underset{\alpha=1,\,6}{\text{med}}\, x_\alpha$, $y^o = \underset{\alpha=1,\,6}{\text{med}}\, y_\alpha$, $z^o = \underset{\alpha=1,\,6}{\text{med}}\, z_\alpha$. For the example, the median guess solution is $x^o = -45147.9$ ft., $y^o = 87423.8$ ft., $z^o = 11117.3$ ft. After five iterations the sequence has converged to the solution $x = 32964.8$ ft., $y = 87425.2$ ft., $z = 11114.9$ ft. The residuals from the final solution are

Residuals

Cine	Azimuth	Elevation
1	.000008	-.000064
2	-.242553	.011513
3	.000022	.000081
4	.000057	-.000019

Thus, the robust solution using the Hampel ψ with breakpoints of 3, 6, 9, correctly identified the outliers. Let us carry this example farther. Suppose we have no observations from Cine 1, i.e., we have data from only three cines one of which is bad. In this case our starting solution turns out to be $x^o = 45147.9$, $y^o = 87424.1$, $z^o = 11120.2$. After four iterations the solution has converged to $x = -32966$, $y = 87424.6$, $z = 11115.3$. Thus, we are again able to correctly identify the bad cine. Now suppose we have data from cines 1, 2, 3. In this case the initial guess solution is $x^o = 45147.9$, $y^o = 67033.9$, $z^o = 11118.9$. After ten iterations the solution is $x = -35023.9$, $y = 84462.1$, $z = 11004.1$. The solution eventually converges to the correct value, but slowly. A third possibility to have data from only three cines is obervations from cines 1, 2, 4. In this case the guess solution is $x^o = -46454.3$, $y^o = 87548.3$, $z^o = 7262.7$. After three iterations the solution has converged to $x = -35392.6$, $y = 86464.3$, $z = 1044.8$. Thus, in this case the iteration has converged to the wrong solution. In the last two cases where the solution converged very slowly and converged to the wrong solution, the starting solution was too far from the correct solution. If a sufficiently good start had been provided the solution would have converged correctly in a few iterations. If the number of cines were great enough in comparison to the number of bad cines, using the median of the solutions obtained from the cine pairs provides an acceptable starting solution. Unfortunately, the number of cines is often no more than three or four. In the case of three cines the use of a starting solution predicted from preceding points might be a desirable procedure. If preprocessing had been used on the cine data most if not all of the outliers of the spike variety in the cine data would have been detected before attempting a solution. Thus, robust estimation in the solution has only to contend with detecting badly biased cines. In any situation with three or more cines with one bad cine, the robust solution will usually provide a better solution than the usual least square procedure. A strategy for choosing a good starting solution needs to be developed. A robust N-station radar solution is developed along the same lines as a robust cine solution. In the radar case a starting solution for the iteration is somewhat easier to obtain.

Application to Recursive Filtering

Very little development has been done on the application of robust statistical techniques to filtering. The most significant effort known to the author is given in the paper of Masreliez and Martin [7]. Their development of robustifying the Kalman filter is quite complex and will not be considered here. It is a simple matter to specify a form for an approximate M-filter and its covariance.

Suppose we wish to estimate the state x(n) of the linear dynamic model described by the state equation

$$x(n + 1) = \Phi(n + 1, n)x(n) + u(n) \tag{37}$$

where $\Phi(n + 1, n)$ is an mxm state transition matrix and u(n) is an m-vector of state noise with covariance Qn. Suppose we are also given scalar observations Z(n) of the state specified by

$$Z(n) = Hx(n) + v(n) \tag{38}$$

where H is a 1xm matrix of constants and v(n) is observation error. By analogy with the least squares filter derivation we minimize

$$\sum_{i=1}^{N} \rho\left(\frac{Z(i) - Hx(i)}{s_i}\right) + \frac{1}{2} u(i)Q_i^{-1}u(i) \tag{39}$$

Subject to the constraints

$$x(i + 1) - \Phi x(i) - u_i = 0 \quad i = 1, n - 1$$

Minimizing (39) leads to the approximate filter equations

$$\hat{x}(n + 1/n + 1) = \hat{x}(n + 1/n) + \frac{P_{n + 1}H^T}{s_{n + 1}} \psi\left(\frac{Z(n + 1) - H\hat{x}(n + 1/n)}{s_{n + 1}}\right) \tag{40}$$

$$\hat{x}(n + 1/n) = \Phi\hat{x}(n) \tag{41}$$

with approximate covariance

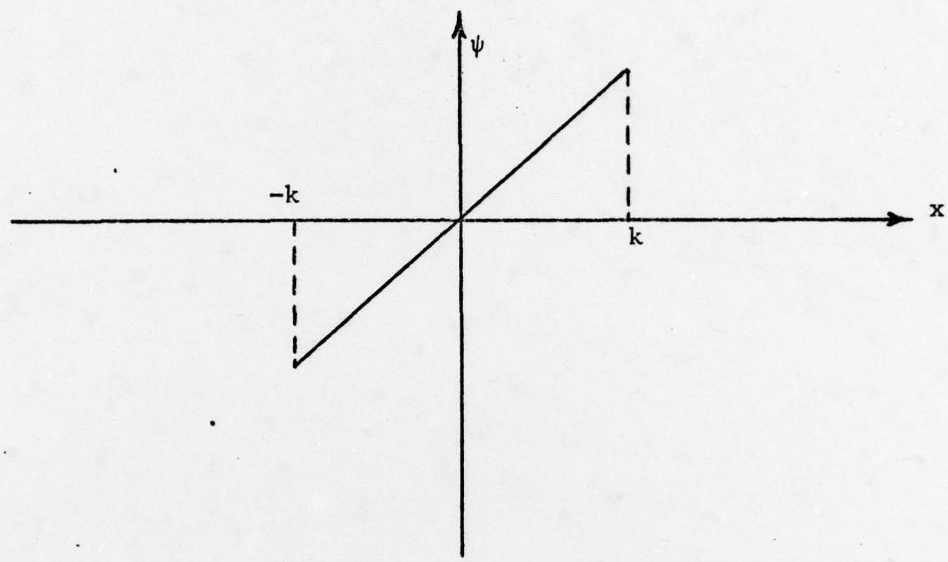
$$P^{-1}_{n + 1} = P^{-1}_{n + 1/n} + \frac{H^T H}{s_n^2} \psi'\left(\frac{Z(n + 1) - H\hat{x}(n + 1/n)}{s_{n + 1}}\right) \tag{42}$$

$$P_{n + 1/n} = \Phi P_n \Phi^T + Q_n \tag{43}$$

Where $\psi(\cdot)$ is an appropriate influence function. Note that the derivative of ψ is required for the update of the filter covariance matrix.

This robust filter is certainly easy to implement and anyone who has done much recursive filtering of data on a computer has probably implemented such a filter with the following ψ function,

$$\psi(x) = \begin{cases} x & |x| \leq k \\ 0 & |x| \geq k \end{cases}$$



i.e., we process observations only if the predicted residuals are within $\pm k\hat{\sigma}$ where $\hat{\sigma}$ is an estimate of the measurement noise standard deviation. Thus, robust filtering presents nothing new as far as filter implementation is concerned, but we are now in a position to possibly improve our robust filtering by borrowing some ψ functions and other concepts which have proved very useful in robust regression.

REFERENCES

1. Huber, Peter J., "Robust Regression: Asymptotics, Conjectures, and Monte Carlo", Annals of Statistics, 1, (1973), pgs. 799-821.

2. Huber, Peter J., "Robust Statistics: A Review", Annals of Mathematical Statistics, 43, (1972) pgs. 1041-1067.

3. Tukey, John W., and Beaton, Albert E., "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data", Technometrics, 16, (May 1974), pgs. 147-192.

4. Ramsay, J. O., "A Comparative Study of Several Robust Estimates of Slope, Intercept, and Scale in Linear Regression", Journal of the American Statistical Association, 72, (Sept. 1977) pgs. 608-615.

5. Theil, H., "A Rank-Invariant Method of Linear and Polynomial Regression Analysis", Indag. Math., 12, (1950), pgs. 85-91, 173-177, 467-482.

6. Hampel, Frank R., "The Influence Curve and its role in Robust Estimation", Journal of the American Statistical Association, 69, (June 1974), pgs. 383-393.

7. Masriliez, Johan C., and Martin, R. Douglas, "Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter", IEEE Transactions on Automatic Control, AC-22, (June 1977), pgs. 361-371.

8. Andrews, D. F., "A Robust Method for Multiple Linear Regression", Technometrics, 16, (Nov. 1974), pgs. 523-531.

9. Mood, A. M., "Introduction to The Theory of Statistics", McGraw-Hill, New York, 1950.

10. Fletcher, K., and Powell, M.J.D., "A Rapidly Convergent Descent Method for Minimization", Computer Journal, 6, (1963), pgs. 163-168.

11. Agee, W. S., and Turner, R. H. "Robust Regression: Some New Methods and Improvements of Old Methods", Technical Report, White Sands Missile Range, 1978.

12. Agee, W. S., and Turner, R. H. "Robust Regression: Computational Methods for M-Estimates", Technical Report, White Sands Missile Range, 1978.